

Data	Source	Data year	Use	Modification	Assumptions for missing data
Country names, codes and boundaries	https://gadm.org/	version 3.6	To plot the shapefile on top of the maps and to create a gridfile that links each grid to a subarea.	The boundaries have been used to develop isogrid files that for each cell provide the country it is inside. Because R cannot deal with factors, we made up our own numbering system. The isorasters are therefore matched with a .csv file that links iso3 codes, (sub)country names and the code in the isoraster file. We created isorasters at 0.5 degree resolution for level 0 input data and 0.1 degree resolution for level 1 input data. We also created an isoraster at 0.0083333 degree resolution for level 0 input data. The boundary shapefile is used to plot the boundaries on top of the gridded results.	There are no missing data.
Population and fraction urban	https://population.un.org/wup/Download/	2018	The emissions are dependent on the population size, as in an area with higher population density an		Countries without data will not have output.
Fraction of population under 5 years of age	https://population.un.org/wpp/Download/Standard/Population/	2020	For some pathogens the incidence is different for younger children compared to the rest of the population	The fraction <5 is calculated by dividing the population in the category under 5 by the total population.	Fractions under 5 are copied from similar countries. See next tab for details.
HDI	http://hdr.undp.org/en/data	2018	The incidence is lower in areas with an HDI higher than the HDI boundary, compared to areas with lower HDI.	We take the cut-off point 0.800: The cut-off points are: HDI of less than 0.550 for low human development, 0.550-0.699 for medium human development, 0.700-0.799 for high human development and 0.800 or greater for very high human development.	HDI's are copied from similar countries. See next tab for details
HDI boundary	http://hdr.undp.org/en/content/human-development-report-2019-readers-guide	2018	We use this boundary to determine the incidence in the population	We create urban and rural population files from this gridded total population file. For each country we rank the population grids from high to low. The grids with highest population become urban, until the total urban population (calculated using the urban fractions) is reached. The remaining grids become rural grids. We first aggregate the population over a 0.5 degree resolution and then divert urban and rural population.	There is no missing data
Gridded population	https://landscan.ornl.gov/	2014	We distribute subarea outputs over the urban and rural populations in the subareas. Grids with higher population density have more emissions		There is no missing data
Toilet category fractions	https://washdata.org/data/download/WLD		The way the pathogens flow through the sanitation chain and get released to the environment most recent, si differs from one toilet category to the next. See further explanation below.	For each country the most recent and complete data on the toilet categories (from the household excel file) are selected. For the country dataset the country files are used. For the subnational dataset, the country files are used together with the inequalities files.	NA values are provided for missing data
Wastewater and fecal sludge management	https://washdata.org/data/household#/table?geo0=country	2017	The way the waste is managed determines the pathogen removal and release into the environment is explained below.	We download sanitation data for all 229 countries analyzed by safely managed criteria	For sewage treatment, NA values were replaced with regression: $\ln(\text{sewageTreated}) = 8.7295 * \text{HDI} - 8.0146$ sewageTreated was assumed to be 0.22 if HDI was not available. For fecal sludge management in countries with HDI < 0.8, NA values in fecalSludgeTreated were replaced with 0.2 (urban) or 0 (rural), coverBury values were replaced with 0 (urban) or 0.9 (rural). For fecal sludge management in countries with HDI > 0.8, NA values in fecalSludgeTreated were replaced with 0.9 (urban) or 0 (rural) and coverBury values were replaced with 0 (urban and rural).
Percent removal by wastewater treatment	https://ars.els-cdn.com/content/image/1-s2.0-S0301479718311824-mmct1.pdf	2010	The wastewater treatment removes the pathogens from the liquid part of the waste before the effluent is released into the environment. The pathogens in the solid fraction of the waste are not included in the model and not released into the environment.	Primary, secondary and tertiary fractions are used together with the log removal and percentage in liquid from sketches on ordinary primary, secondary and tertiary treatment systems.	
Incidence, shedding rate and shedding duration for	As explained in: Hofstra et al 2013 for Cryptosporidium and Kiulia et al 2015 for rotav		The incidence, shedding rate and shedding duration together determine the excretion per person in the subarea.		There is no missing data, assumptions are only based on HDI.